Available online at www.sciencedirect.com
SCIENCE DIRECT®
ACADEMIC PRESS
Organizational Behavior and Human Decision Processes 90 (2003) 124–138
ORGANIZATIONAL BEHAVIOR AND HUMAN DECISION PROCESSES
www.elsevier.com/locate/obhdp

Precision of probability information and prominence of outcomes: A description and evaluation of decisions under uncertainty

Johannes Hönekopp*

Technische Universität Chemnitz, Psychologisches Institut, Wilhelm-Raabe-Str. 43, 09120 Chemnitz, Germany

Abstract

Information about event probability upon which decisions depend may be more or less precise. The first section of this paper reports three experiments that investigated the relationship between this type of imprecision and the prominence that outcomes obtain in decisions. Participants had to rank order sets of six lotteries according to attractiveness. While the lotteries' values were always precisely known precision of information about lottery chances varied. These experiments showed that increasing ambiguity tied decisions closer to lottery values. The second section shows that modeling participants' decisions with the contingent weighting model suggests that this outcome prominence effect was not necessarily caused by any change in the respective weighting of probability and outcome information, but that it had probably occurred for purely mathematical reasons. The third part of this paper explores, by means of a computer simulation, (i) which weighting strategy is optimal when probabilities are imprecise and (ii) how participants' decision behavior compared to a simple, but better adapted strategy. It shows that the weighting of probability information should not change with decreasing precision and it implies that participants' performance suffered most from a lack of strategic consequence. Implications for decision making policy in general are discussed.

Keywords: Decision making under uncertainty; Weighting of probabilities; Prominence of outcomes; Imprecise probabilities; Improving decision making

1. Introduction

Most theories that seek to describe decision making under uncertainty assume that the attractiveness of each option is closely related to its subjectively expected utility, that is, possible outcomes (or derivatives of these) and their subjective probabilities (or derivatives of these) are combined multiplicatively. Camerer and Weber (1992) give an overview over such models in the wake of Savage's (1954) subjectively expected utility theory. Very often outcome probabilities are not exactly known but are rather imprecise. If you bet on a coin toss you would probably assume your chance of winning to be .5; however, if you bet on the upshot of a horse race you might feel reluctant to state your chances of winning as a precise number. Or consider a patient who ponders whether to undergo surgery or not. If she is told that it has been performed 5000 times, 4000 times successfully thereof, she might feel that .8 is the "true" chance of success. But if she is informed that this treatment has been tried only five times, four times successfully and once unsuccessfully, she might think different about her chances. Regarding any probability estimate from such a small sample as quite uncertain she might for instance, just to be on the safe side, assume her "true" success probability to be only .5 (Weber, 1994).

One common source of imprecision of probability information or ambiguity, as it is often called, stems from a widespread preference (Erev & Cohen, 1990; Wallsten, Budescu, Zwick, & Kemp, 1993) to inform others about probabilities or degrees of belief by using verbal descriptions like *very likely*, *improbable* and so on instead of clear cut numbers. Such probability descriptions are known to have no fixed meaning (people do not think that *very likely* denotes a probability of .92 or

* Fax: +49-371-531-63-33.
E-mail address: johannes.hoenekopp@phil.tu-chemnitz.de.

0749-5978/03/$ - see front matter © 2003 Elsevier Science (USA). All rights reserved.
doi:10.1016/S0749-5978(02)00514-9

.87). The meaning of such phrases can be more adequately captured either by a probability range (e.g., Mosteller & Youtz, 1990) or by a membership function that tells how much any probability is warranted by the meaning of a particular phrase (Wallsten, Budescu, Rapoport, Zwick, & Forsyth, 1986). That is, phrases that communicate a degree of uncertainty are inherently ambiguous.

González-Vallejo, Erev, and Wallsten (1994) have shown that the respective contribution of probability and outcome information to decisions depends on whether decision makers (*dms*) are given verbal or numerical probability information: In this study, participants were confronted with sets of six lotteries, which had to be rank ordered. Each lottery was played once and would either win with probability p or loose. If a lottery won, it yielded its worth \$ multiplied by its assigned rank, otherwise nothing happened. Thus, dms were motivated to assign rank six to the best lottery, rank five to the second best, etc. Each set of lotteries had to be rank ordered twice. In one trial the probabilities p were stated numerically. In the other trial, verbal descriptions of these probabilities (e.g., *improbable*), which were provided by other participants, were given instead. González-Vallejo et al. (1994) investigated whether the type of probability information (numerical or verbal) had an effect upon the correlation between the lotteries' ranking and the lotteries' values \$. They found dms' rankings to be tied more closely to the lotteries' values in the verbal probability mode.

It seems plausible that the ambiguity of verbal probability phrases caused the findings of González-Vallejo et al. (1994). And it seems likely that ambiguity achieved by other means will have a similar effect. For the sake of simplicity, I shall call the presumed effect that imprecision of probability information causes decisions to be tied more closely to lottery values *outcome prominence effect*.

The first aim of this paper is to explore the assumed outcome prominence effect. Does ambiguity indeed enhance the prominence of options' outcomes? If so, is the relationship between the degree of imprecision and the outcome prominence effect's strength linear? Three experiments in which the precision of probability information was systematically varied will be reported to respond to these first two questions (Section 2). The second aim of this paper is to better understand what causes the outcome prominence effect (Section 3). The last aim is to identify which response to ambiguity is optimal and to explore the consequences of participants' decision making behavior. To do so, the latter will be emulated and embedded into a computer simulation that tests participants' behavior under a broad variety of circumstances versus a simple but more reasonable strategy (Section 4).

2. The impact of probability informations precision on outcome prominence

All three experiments reported here share a basic paradigm, which is closely related to the experimental approach of González-Vallejo et al. (1994). This paradigm is described in the next paragraphs.

2.1. The Basic paradigm

As in González-Vallejo et al.'s (1994) experiment, participants had to deal with two-outcome lotteries, either of the form "win point value \$ with probability p, or else nothing" (gains) or "lose point value \$ with probability p, else nothing" (losses).[1] The lottery values were randomly drawn from the arbitrarily chosen interval [3, 97], and the probabilities from the interval [.03, .97]. Both, \$ and p, were uniformly distributed and uncorrelated. This ensured that both dimensions, \$ and p, contributed equally to the variance of the lotteries' expected values. The lottery values were numerically provided, and the probabilities were presented by means of spinners. Such a graphical representation of probability is not uncommon, and studies suggest that they produce quite similar results as those achieved with numerical representations (Budescu & Weiss, 1987; Wallsten, 1971). Six lotteries together formed a set (see Fig. 1); these had to be rank ordered and the same mechanism of outpayment described above was applied.

Imprecision of probability information was obtained by concealing a part of the spinner with an occlusion; the occlusion size defined the degree of ambiguity and will be reported as percentages. The occlusion was randomly positioned with two restrictions; that (i) exactly one of the two boundaries between the spinner's winning and neutral area was covered, (ii) at least 1% winning area and also at least 1% neutral area lay below the occlusion (see Fig. 2). Thus the size of the winning area (i.e., p) could not be exactly determined. All lotteries in a set were presented with the same degree of ambiguity. In the remainder of this text, I will refer to the visible part of the winning area as p_{visible} and to the size of the occlusion (degree of ambiguity) as *ambig*.

Note that the restrictions for the placement of the occlusion are different for different p values: For example, if p is .5 and *ambig* is .4, p_{visible} can take values between .11 and .49. But if p is .1, p_{visible} can only vary within a much narrower range, between .01 and .09. The restrictions bring along that although p is uniformly distributed $(p|p_{\text{visible}})$ is not. Consequently, the best estimate for p is not $(p_{\text{visible}} + ambig/2)$, as one might intuitively think, but an extremer value, i.e., a value spread

[1] To keep things simple, I will consider only gains in the following description. Of course, all details hold for losses as well.

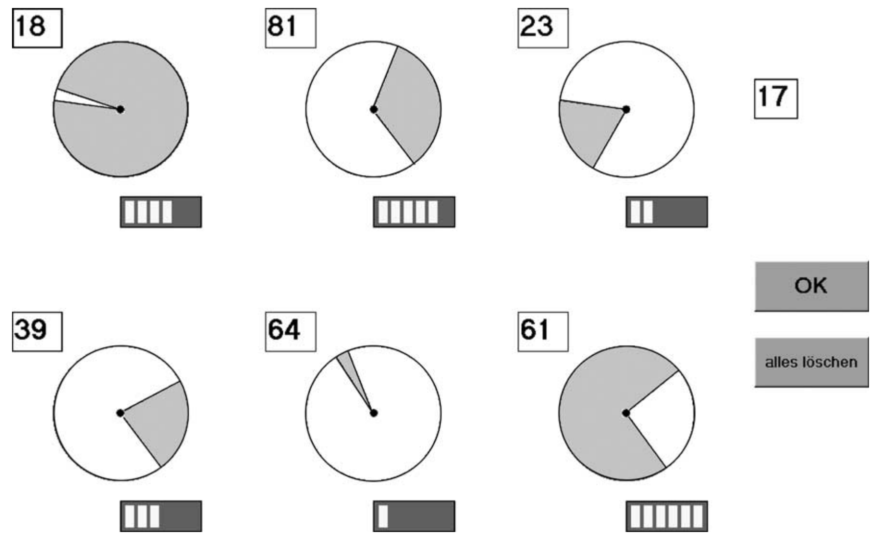

Fig. 1. Screen with a set of six lotteries. Spinners' dark areas indicate the chance to win the lottery. Lottery values are indicated in the boxes top left from the spinners. Valuation factors are already assigned (bars below the lotteries; e.g., factor 4 for the top left lottery). The single box to the right indicates the number of sets still to be played (here 17).

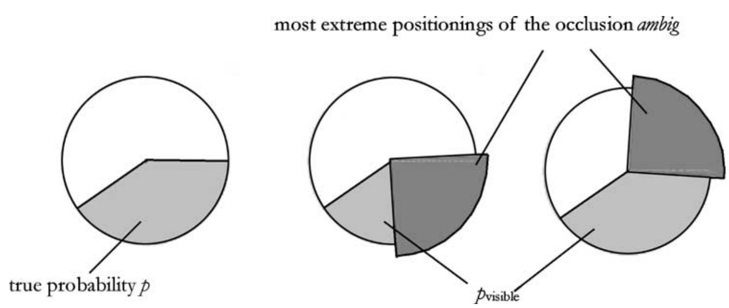

Fig. 2. Construction of ambiguous lotteries. In a first step, *p* was defined. Then the position of the occlusion (here 25%) was randomly determined within the stated restrictions. The two most extreme positions possible are depicted. Of course, experimental subjects could not see where the area that indicates *p* ends.

away from .5. The optimal strategy for resolving ambiguity is depicted in Hönekopp's (2000) Appendix.

2.2. Experiment I

The purpose of the first experiment was to demonstrate the outcome prominence effect and to explore whether the relationship between degree of ambiguity and outcome prominence is linear. A broad range of ambiguities was employed to address this question.

Method. 26 women and 32 men participated in the experiment; median age was 23 years old. Subjects, most of them students, were recruited via fliers in university cafeterias and ads in a small local magazine. They were paid € 2.5 for participating, psychology undergraduates could alternatively fulfill research requirements. All participants competed for premiums of € 40, 30, and 20, which were promised to the skillful three who would gain the most points.

The experimental procedure was as described above. The experiment involved a within subjects design. The independent variable, degree of ambiguity, had four levels: 0%, 25%, 55%, and 85%. Each participant assigned valuation factors to 60 sets of lotteries, 15 of each ambiguity level. Sets were presented in blocks, and block orders were individually randomized. For each participant, all lotteries were randomly composed, following the guidelines sketched above.

The lottery sets were administered on a computer screen (Fig. 1). The valuation factors were assigned in descending order by clicking the small gray boxes below the spinners. The first lottery selected received factor six, the second factor five, and so on. Of course dms could correct their choices. Pressing the "ok" button completed a set of lotteries, and the next set then appeared on the screen. Additionally the small counter in the upper right hand corner (showing 17 in Fig. 1), which informed dms about the number of sets still to be ranked, decreased by one. Dms could neither perceive which lotteries in a set had won, nor the amount of points they had gained. They could merely see their final result at the end of the experiment. Different from what the dms were told, lotteries were not actually played, instead participants received, for each lottery, its expected value times its valuation factor. This was done in order to avoid random error in the performance data.

Participants first read the printed instructions. This informed them in detail about the way in which the positioning of the occlusions was determined. Thus, dms could adopt the optimal strategy at least hypothetically. When participants had finished reading the instructions, the experimenter demonstrated the computer program and explained the task once more. This took place in groups of up to four people. Participants then worked alone or two to a room. Before starting with the "real" lotteries, participants were free to complete as many practice trials as they chose. Practice trials consisted of lotteries without ambiguity and provided no feedback. Dms worked at their own pace and took an average of about 30 min to complete the 60 sets.

Results. In a first step performance data were analyzed to check whether participants had understood the task. If a dm does not comprehend the task and arbitrarily assigns valuation factors to the lotteries his rankings will be uncorrelated with the lotteries' values, as well as with their probabilities. I made use of this fact to sort out participants without task understanding: For each dm and each set, a performance measure $[V(rdm, r\$) + V(rdm, rp)]/2$ was computed, where V is a rank correlation measure that indicates a perfectly positive correlation with 1, a perfectly negative correlation with 0, and a lack of coherency with .5 (see Nelson, 1984, for details). *rdm* indicates the dm's ranking, *r*\$ signifies the ranking that occurs when lotteries are ordered according to their values, and correspondingly, *rp* denotes that ranking that is obtained by ordering lotteries according to their probabilities. For each dm, all 60 performance values were averaged. A scatterplot showed two clear outliers with averages <.55 (remember that the expected value for random ranking is .5). The median for the other dms was .72. The data of these two participants were excluded from all further analyses.

For each dm and each lottery set, $V(rdm, r\$)$ was computed to measure the association between the ranking of the lotteries and their outcomes, i.e., the outcome prominence. For each dm and each experimental condition, all fifteen *V*-correlations were averaged; the following analysis draws upon these averaged values. A graphical examination of the data distributions justified a parametric analysis. As can be seen from Fig. 3, increased imprecision of probability information always caused an increasing prominence of lottery values \$. The overall pattern could be well captured by a linear contrast, whose weights mirrored the respective differences in ambiguity: The correlation between the group means and their corresponding weights was $r_{\text{alert}} = .99$ (see Rosenthal, Rosnow, & Rubin, 1999). The linear contrast proved to be statistically significant

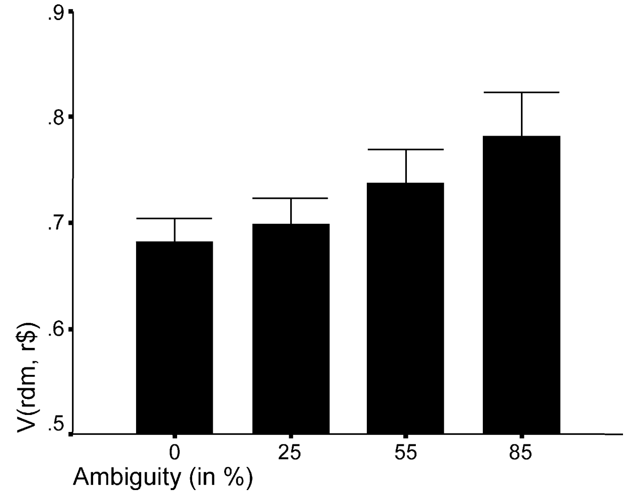

Fig. 3. Average *V*-correlations between dms' ranking of the lotteries and the one that is obtained by ranking lotteries due to \$; Experiment I. Flankers indicate the upper ends of the .95 confidence intervals.

with $t(55) = 5.6$ ($p < .001$). Following the suggestion of Rosenthal et al. (1999), I computed the effect size of the linear contrast as if a between subjects design had been employed, which resulted in $r_{\text{effect size}} = .31$.

Discussion. The results showed a marked outcome prominence effect. Furthermore, the relationship between degree of ambiguity and outcome prominence proved to be highly linear. Interestingly, in the domain of gains the outcome prominence effect leads to decisions that are increasingly risk seeking. Imagine the choice between the following two alternatives: (a) win \$40 with $p = .5$, or else nothing and (b) win \$50 with $p = .4$, or else nothing. The larger the impact of outcomes is the likelier the riskier option (b) will be chosen. In the absence of ambiguity, people tend to be risk averse in the domain of gains (Kahneman & Tversky, 1979) and this was true for the dms in Experiment I as well: In the experimental condition of precise information, $V(rdm, rp)$ (with $M = .80$) clearly exceeded $V(rdm, r\$)$ ($t(55) = 6.3, p < .001$). Thus, with rising ambiguity, decisions became increasingly risk seeking, although dms exhibited a strong preference for safer options in the absence of ambiguity.

What can be expected for losses? People tend to be risk seeking in this domain (Kahneman & Tversky, 1979). Additionally, an outcome prominence effect would make choices more conservative. Imagine the two alternatives from the previous paragraph, but now with negative outcomes. Increasing prominence of the outcomes will change the odds in favour of alternative (a) being chosen, which is the less riskier of the two. Thus, in a way, an outcome prominence effect has the same consequences in both domains: It changes decisions, contrary to the initial risk preferences. Therefore, one may expect that ambiguity has similar effects on the prominence of outcomes in both domains. The aim of the next experiment was to replicate the previous finding and to widen the scope to the domain of losses.

2.3. Experiment II

Method. 42 women and 18 men, hired by the same means as before, participated in Experiment II. Again, most of them were students. Participants were paid € 6 each, psychology undergraduates could alternatively fulfill research participation requirements. All participants competed for premiums of € 40, 30, and 20 that would be attributed to the best three. Experiment II employed a 2×4 within subjects design, with the factors domain (gains vs. losses) and ambiguity (0%, 25%, 55%, and 85%). Each of the eight resulting cells consisted of 10 sets of lotteries. Construction and presentation of the tasks followed the same guidelines as in Experiment I. In the domain of losses a subtraction sign marked lottery values and additionally *loss loss loss* was written in red on the middle of the screen. For negative gambles, the already known rules held too, that is, dms were told that if such a lottery lost, its point value \$ times the valuation factor that had been assigned to this gamble would be subtracted from their assets, otherwise nothing would happen. But in effect each lottery's expected value times its factor was accounted for the aforementioned reason.

The procedure differed from Experiment I by one detail: Up to four dms worked together in a room. Computers were arranged in such a way that no participant could look at the screen of any other dm. Unbeknownst to the participants, decision time was tracked automatically for every set.

Results. Due to a software error the ranking of the last lottery set was lost for 15 participants. For the same reasons the decision times for the last five sets were lost for 18 participants. Losses were unsystematically spread over the eight experimental conditions and the data analyses could be performed as planned, although sometimes based on fewer sets.

Whether participants had understood the task was checked as before in Experiment I. Five extreme outliers with averaged V-correlations <.55 were excluded from all further analyses. A closer inspection of the remaining data showed for a few participants in single conditions, averaged V-correlations as low as .22, i.e., considerably below the chance level of .5. Much evidence suggested that in these cases, dms failed to notice a change either from a block with winning sets to a block with negative lotteries or vice versa, and thus acted systematically against their own interest. In this case, which pertained to eight from a total 440 data blocks (eight experimental conditions times 55 dms), the actual data were discarded and replaced with the corresponding sample medians.

Fig. 4 shows the mean $V(rdm, r\$)$ values for all experimental conditions. The pattern in the positive domain largely resembles the one found in Experiment I. As

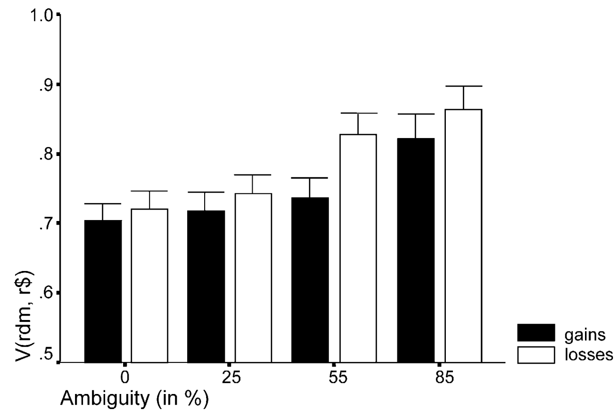

Fig. 4. Average V-correlations between dms' ranking of the lotteries and the one that is obtained by ranking lotteries due to \$; Experiment II. Flankers indicate the upper ends of the .95 confidence intervals.

can be seen, ambiguity produced a monotonous outcome prominence effect in the negative domain, too. And at all ambiguity levels, $V(rdm, r\$)$ was somewhat higher for losses than for gains. The influence of imprecision could be well described by the same linear contrast employed in Experiment I. After correcting the cell means for the effect of domain these correlated very highly again with the contrast weights ($r_{\text{alert}} = .95$), i.e., the influence of ambiguity on $V(rdm, r\$)$ was highly linear in both domains. The linear contrast was statistically significant ($t(54) = 8.4$, $p < .001$). As Fig. 4 shows, there was no substantial interaction between the two manipulations. Therefore, the effect of the level of imprecision can be described best when domain is treated as a non-substantive factor and its influence is partialed out (Rosenthal et al., 1999, p. 80ff). It proved $r_{\text{effectsize|NS}} = .15$ (again, this effect size was computed as if a between subjects design had been employed). The influence of domain was inspected by a second contrast analysis where all four cells of the positive domain were weighted by +1 and the other cells by −1. These contrasts well captured the influence of the domain ($r_{\text{alert}} = .83$, with effect of level of ambiguity partialed out) and proved to be significant ($t(54) = 3.7$, $p < .001$). Domain was associated with an effect size of $r_{\text{effect size|NS}} = .17$ (this time ambiguity was treated as a non-substantive factor and, again, the effect size was computed as if a between subjects design had been employed).

Discussion. Experiment II replicated the finding of the first study. In addition, a similar outcome prominence effect was found for losses and the effect of ambiguity could be well described across domains with the same set of weights. As described above, the outcome prominence effect in the negative domain signifies that decisions became increasingly conservative.

One may criticize the two experiments because ambiguity was achieved in a rather artificial way, and one may argue that a very different picture would have emerged if imprecision of probability information had been realized in a more natural way. Experiment III set out to check this possibility. It compared the effects of ambiguity from two different causes: The already known one, owing to partly occluded spinners, as well as verbal descriptions of probabilities. As already mentioned, such verbal descriptions are inherently ambiguous and frequently used. Therefore, verbal probability descriptions can be seen as a common source of ambiguity in natural situations.

2.4. Experiment III

Verbal translations. Experiment III used the same task as the previous two experiments. Since a verbal description had to be obtained for each probability used, lotteries were not constructed anew for each dm. Instead, only 60 different lotteries were used, which formed 10 sets, and thus comparably few verbal translations of probabilities were needed. The lottery sets were chosen as follows: Several sets were randomly constructed following the guidelines described above. From these, 10 sets, which are provided in Hönekopp's (2000) Appendix, were selected so that the subsequent restrictions were fulfilled: (i) The average of p was close to .5 and the average of \$ close to 50. (ii) \$ and p showed variances similar to those of the foregoing experiments. (iii) For each set, two rank correlations $V(rev, r\$)$ and $V(rev, rp)$ can be computed, where rev is the rank order that is obtained when lotteries are ranked according to their expected values; the averages of $V(rev, r\$)$ and $V(rev, rp)$ were to be similar. That is, on average lotteries' values and lotteries' true probabilities carried equal amounts information about the lotteries' expected values. The 60 gambles happened to consist of 46 different probabilities.

To gather verbal descriptions for these, the probabilities were represented as spinners and printed out in random order, six at a time, on a sheet of paper. Three colleagues without knowledge about the aim of the experiment, provided verbal descriptions of these chances, thereby taking into account the following restrictions: (i) No numbers were allowed, with the exception of *50–50*. (ii) Descriptions were not allowed to refer to a clock face, thus forbidding *short past seven o'clock* and the like. (iii) For practical reasons, phrases had to be rather short. (iv) All descriptions had to suit positive as well as negative consequences, i.e., all valuing terms like *luck*, *hope*, *danger* and the like were prohibited. (v) All formulations were to be understood as probabilities even without context, thus *rather small probability* was appropriate while *rather small* was not. Last but not least the three translators were ensured that there was no need for abundant variety in formulations. They used between 6 and 18 different phrases for the 46 probabilities.

For a meaningful comparison of the effect of the verbal descriptions with the effect of the spinners, it is necessary to match them with respect to degree of ambiguity. Therefore, the next step aimed at estimating the degree of ambiguity inherent to the verbal translations. To do so, all of the three translators' phrases were mixed together into a single list, which was then given to five student raters. It was explained to these students how the phrases had come into being, and they were asked to provide, for each single phrase, a probability range representing its possible meaning by indicating an upper and lower bound. As such, for each of the three original lists, the average size of these confidence intervals could be computed. In this respect, the original lists of the three translators differed only insignificantly; the overall average range size was 12.5%. I know of only one study that reports average sizes of ranges of meaning for verbal probability phrases: Reagan, Mosteller, and Youtz (1989) found an average of 12%, which is very much in line with my own data. To match the average degree of ambiguity inherent in the verbal phrases, I decided for an occlusion size of 12% for the spinners.

Method. Experiment III involved a 2×3 within subjects design. Domain (gains vs. losses) was crossed with type of probability presentment: In addition to verbal probability information two controls were used, graphical presentation without ambiguity and, to match the imprecision of the verbal descriptions, graphical presentation with 12% ambiguity. In the verbal condition each of the three lists with probability translations was used for one third of the participants, lists being randomly collated to dms. Again sets were presented block-wise with block order randomly determined for each participant. Each condition consisted of the same ten sets of six lotteries. In each of the 60 resulting sets and for each participant, the positions of the six lotteries in a set were randomly assigned anew, and it was virtually impossible to detect that each lottery occurred six times over the course of the experiment. In the graphical condition with ambiguity the position of the occlusion was randomly determined for each dm for each set.

Forty-three women and 15 men, who were recruited as in the previous experiments, took part in the study. Again most participants were students, median age was 21 years old. They were paid € 4 or optionally fulfilled research requirements. All participants competed for premiums of € 40, 30, and 20, which were promised to the best three.

It was explained to participants how the verbal probability descriptions had been derived. All other details of the experimental proceeding were identical to the second experiment with one exception: To prevent dms from overlooking changes from win to loss blocks or vice versa, these were always accompanied by a short buzz. When participants had finished the 60 sets they were given a questionnaire that contained, in random order, the probability phrases that they had seen in the experiment. Participants were asked to give for each phrase a best estimate of its meaning and, additionally, an upper and a lower bound; I shall refer to the best estimate as $p_{\text{best-guess}}$. One participant did not understand this task; her data were excluded from all further analyses. A last question asked whether dms had felt better informed by the verbal probability descriptions or by the occluded spinners.

The retranslation from phrases to numbers was scheduled at the end of the experiment because Erev, Bornstein, and Wallsten (1993) showed that decisions based on verbal probability descriptions systematically change when dms are required to translate the phrases into numbers beforehand.

2.5. Results and discussion

The data of four dms were excluded from all further analyses for the same reason as in Experiments I and II.

As can be seen from Fig. 5, an unexpected pattern emerged in which ambiguity had no systematic effect on the prominence of lottery outcomes. Therefore, data were analyzed by means of a repeated measures ANOVA. Ambiguity proved to be insignificant ($F(2,51) = .2$, $p = .79$). As in the foregoing studies, the effect of domain was significant ($F(1,52) = 5.5, p = .02$) and in the expected direction but qualified by an unexpected interaction ($F(2,51) = 3.5, p = .03$). In three of the four cases, imprecision of probability information went along with (marginally) increased outcome prominence (see Fig. 5). Reconsidering the results of the previous experiments, ambiguity simply may have been too moderate to bear a detectable effect on $V(rdm, r\$)$ in Experiment III. The interaction effect is difficult to explain. As two post-hoc performed t tests for dependent groups showed, the differences between the two imprecise information groups did not differ statistically significantly in either of the domains (gains: $t(52) = 1.2$, $p = .25$; losses: $t(52) = 1.6, p = .11$), although this type of testing did not correct p values for multiple testing. More importantly, the differences between the verbal and graphical condition were comparably small. The drop of $V(rdm, r\$)$ in the losses/verbal information condition may simply represent a chance result. Although the results of Experiment III are somewhat inconclusive they do not undermine the notion that graphically and verbally achieved imprecision of probability information have comparable effects.

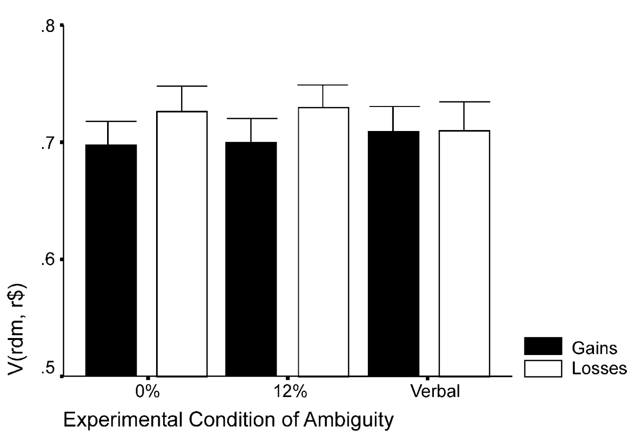

Fig. 5. Average V-correlations between dms' ranking of the lotteries and the one that is obtained by ranking lotteries due to \$; Experiment III. Flankers indicate the upper ends of the .95 confidence intervals.

3. What causes the outcome prominence effect?

By and large, the three reported studies demonstrate a marked linear outcome prominence effect, which is in line with previous research (González-Vallejo et al., 1994). What causes the outcome prominence effect? That dms change their strategy when they face information imprecision and accordingly take the precise information more into account is an idea that comes easily into mind. In line with this notion, Wallsten, Budescu, and Tsao (1997) claimed: "When trading-off among dimensions for the purpose of choosing or evaluating alternatives, the weight accorded a dimension is a positive function of its precision" (p. 32). Applied to the precision of probability information, I shall call this claim *weighting hypothesis*. The results presented here, as well as those of several other studies, point in this direction (Erev & Wallsten, 1993; González-Vallejo et al., 1994; González-Vallejo & Wallsten, 1992; Svenson & Karlsson, 1986). However, this interpretation is not imperative, as noted by the last authors mentioned. Let us consider verbally achieved ambiguity first: Fillenbaum, Wallsten, Cohen, and Cox (1991) showed that receivers of verbal probability descriptions decode them as less extreme (i.e., closer to .5) than as intended by the senders. Thus, it is likely that the verbal descriptions, as represented by the dms, contained less variability than the numerical probabilities. A brief example shows the consequences that arise from this fact: Consider the following lotteries: (a) win \$25 with $p = .8$, or else nothing and (b) win \$80 with $p = .2$, or else nothing. For the sake of simplicity let us assume that a dm's preference order depends on lotteries' perceived expected values; the resulting order (a), (b) is opposite to lotteries' \$ ordering (b), (a). Now assume that lottery probabilities are communicated verbally and are, therefore, understood less extreme as $p = .7$ (a) and $p = .3$ (b). Now, with verbal probability information, lotteries will be ranked (b), (a), i.e., the dm's ranking will now reflect lotteries' \$ ordering. As this example shows, the use of verbal probability information will increase $V(rdm, r\$)$ in the absence of any change in strategy. Therefore, the observed outcome prominence effect cannot tell whether any change in strategy, as implied by the weighting hypothesis, occurred or not. The same will hold for any other type of ambiguous probability information that causes dms' subjective probabilities to regress towards .5.

What causes the outcome prominence effect? Does it originate solely from dms' distorted perceptions of the true probabilities, or is, on top of this, a change of strategy working, as supposed by the weighting hypothesis? In the third part of this paper, I shall try to shed some light on these questions.

How can they be tackled? How can any strategy shift as implied by the weighting hypothesis be described? The contingent weighting model of Tversky, Sattath, and Slovic (1988), which was designed to describe the systematic differences in outcome prominence between choice and bidding tasks (e.g., Lichtenstein & Slovic, 1971), lends itself to this task. The model assumes that lottery (a) in which $\$_a$ is won with probability p_a and otherwise 0, is preferred to lottery (b), in which $\$_b$ is won with probability p_b and else 0, if and only if

$$\$_a p_a^\theta > \$_b p_b^\theta. \quad (1)$$

The more θ exceeds 1, the stronger the relative weight accorded to the lotteries' probabilities; assuming that both gambles have the same expected value, the lottery with the higher chances of winning is preferred. But if $0 < \theta < 1$, more weight is attributed to the outcome dimension than to the probability dimension and the lottery with the higher outcome will be preferred. The weighting hypothesis assumes that dms lower θ with increasing ambiguity.

I modelled the observed decision behavior from the three experiments using the contingent weighting model. θ served as a free parameter that enabled adjustment of the model to the behavior of each dm under each experimental condition. However, the contingent weighting model could not be applied to experimental data in its original form because, in most cases, the dms did not know the lottery probabilities, but had to infer them. Therefore, the estimation of θ had to be based on an assumption about how dms resolved ambiguity. It was assumed that the dms preferred lottery (a) over lottery (b) if and only if

$$\$_a (p_{\text{visible a}} + ambig/2)^\theta > \$_b (p_{\text{visible b}} + ambig/2)^\theta \quad (2)$$

in the case of graphic probability displays and if and only if

$$\$_a p_{\text{best-guess a}}^\theta > \$_b p_{\text{best-guess b}}^\theta \quad (3)$$

in the case of verbal probability information.[2] Unequations (2) and (3) were individually fitted to the empirical data by searching that θ value that obtained the best fit across all sets of a condition for each participant. I employed Kendall's Tau as a measure of fit. The θ values were searched for by a computer program that considered all values between .05 and 50, using a step width of .05. It occurred quite often that not one single, but several adjacent θ values, produced the optimal fit. In this case the lowest value was taken.[3]

3.1. Reanalysis of Experiments I to III

The model fits happened to be heavily skewed to the right. Median values ranged from $\tau = .80$ (Experiment I) to $\tau = .85$ (Experiment III, see also Table 1). To test for the consistency of the θ estimates, the lottery sets were split into odd and even numbers, and θs were estimated in the same way again for each test half. The obtained split-half rank correlations were corrected for test shortening, thereby following the Spearman–Brown formula. The results, which were satisfactory, can be found in Table 1.

Since θ is an exponent all related analyses were performed at an ordinal level. The left of Fig. 6 shows the medians of the θ estimates for the four experimental conditions of Experiment I. As can be seen, rising ambiguity did not cause θ to decrease, as could be expected from the weighting hypothesis. Instead, a tendency in the opposite direction occurred. While a Friedman test did not yield a significant result ($\chi^2 = 3.9; df = 3; p = .27$), a post hoc performed trend test, following Page (1963), which was performed to check for a linear increase of θ, proved to be significant ($p = .026$), indicating that an increase in ambiguity systematically led to a change of strategy, i.e., a more pronounced weighting of lotteries' probabilities.

Why would people base their appraisal of prospects more on probability information when it is less precise? Two explanations seem plausible. Less precise probability information might lead dms to grapple more with lottery chances, as proposed by venture theory (Hogarth & Einhorn, 1990). Thus dms might focus their attention on probability information and, as a consequence, overweight it in their decisions, which is reflected in a rise of θ. I shall call this the *attention hypothesis*. Another explanation might be that people act more cautiously when the information at hand is less precise (as

[2] I used several other modifications of Unequation 1 that differed in the assumption about dms' way of dissolving ambiguity to estimate θ as well. And I also tried models that assumed typical non-linear utility functions for \$. The model that is described in the Unequations (2) and (3) could fit the empirical data best. Models that achieved similar good fits also resulted in similar estimates of θ. A more detailed analysis can be found in Hönekopp (2000).

[3] To test which value is the most commensurable estimate for θ in this case, I used a computer simulation in which artificial dms ranked lotteries, each dm employing a different value for θ. Dms ranked the lotteries following, in general, Unequation 2. But, to represent real participants realistically, their ranking behavior was overshadowed by some random error. Each artificial dm's θ value was then estimated in the described manner. If several adjoining θ values lead to the optimal fit the lowest of these values showed to be closer to the true θ value than the middle or the highest estimate, which were also taken into account.

Table 1
Obtained fit between actual and replicated decisions and consistencies of θ estimates

	Experimental condition	Experiment		
		I	II	III
Split-half consistency of θ estimates				
Gains	0%	.86	.79	.70
	12%			.77
	Verbal			.85
	25%	.86	.73	
	55%	.75	.83	
	85%	.83	.86	
Losses	0%		.82	.76
	12%			.70
	Verbal			.79
	25%		.80	
	55%		.79	
	85%		.79	
Overall fit (τ)	Overall	.80	.84	.85

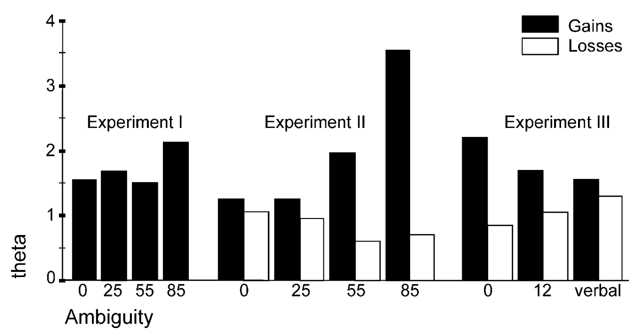

Fig. 6. Median θ estimates for Experiments I–III.

already mentioned, the higher θ rises, the more cautious is behavior in the domain of gains). I will refer to this possibility as *conservatism hypothesis*. Of course, both forces, attention shift and change of attitude towards risk, may go hand in hand.

The setup of the second experiment enables testing of both explanations against each other. If ambiguity causes attention to be drawn to probabilities this should show in the domain of losses as well. Thus, θ should rise along with ambiguity here too. On the contrary, if ambiguity prompts more cautious behavior, θ should increasingly decrease in the domain of losses. The reason is that a lowered θ gives more importance to lottery outcomes which, in the domain of losses, favors conservative bets.

As the middle of Fig. 6 shows, a pattern occurred that clearly favors the conservatism hypothesis. More clearly than in the previous experiment, θ increased along with ambiguity in the positive domain, but the contrary tendency could be observed for losses. Two separate trend tests, following Page (1963), confirmed that the observed patterns were statistically significant ($p < .001$, gains, and $p = .004$, losses).

In Experiment II decision times had been recorded. The attention hypothesis (and venture theory's claim that higher ambiguity leads to higher mental occupation with prospects' probabilities) would receive support if decision times rose jointly with ambiguity. With the conservatism hypothesis, there is no reason to assume any effect from ambiguity on decision speed. To eliminate the effect of outliers the median decision time of each dm under each condition was entered into the analysis. All eight distributions of medians approximately followed a normal distribution. A two way ANOVA with the repeated measurement factors domain (win/lose) and ambiguity (0%, 25%, 55%, and 85%) was performed. Only the domain variable showed a significant effect ($F(1,54) = 44.4,\ p < .001$). Average decision time was 24.0s for gains; in the domain of losses, dms needed on average 19% more time (i.e., 4.6 s). Prolonged decision times for losses are a common finding (e.g., Budescu, Weinberg, & Wallsten, 1988). Neither the factor ambiguity showed any effect ($F(3,52) = 1.4$, $p = .27$) nor its interaction with the domain ($F(3,52) = .6, p = .65$). Therefore, not only the θ pattern supported the conservatism hypothesis, but decision times as well. They challenged not only the attention hypothesis but also venture theory's claim that heightened imprecision of probability information should lead to more extensive processing of the latter. The data showed no tendency in the proposed direction (see Fig. 7).

The θ estimates for Experiment III showed a different pattern than was found previously (see right of Fig. 6); for gains θ decreased with ambiguity (as proposed by the weighting hypotheses), while the opposite held true for losses. While there was some statistically significant effect in the positive domain ($\chi^2 = 6.4, df = 2, p = .04$; Friedman test), the differences in the negative domain were not significant ($\chi^2 = 1.3, df = 2, p = .54$; Friedman

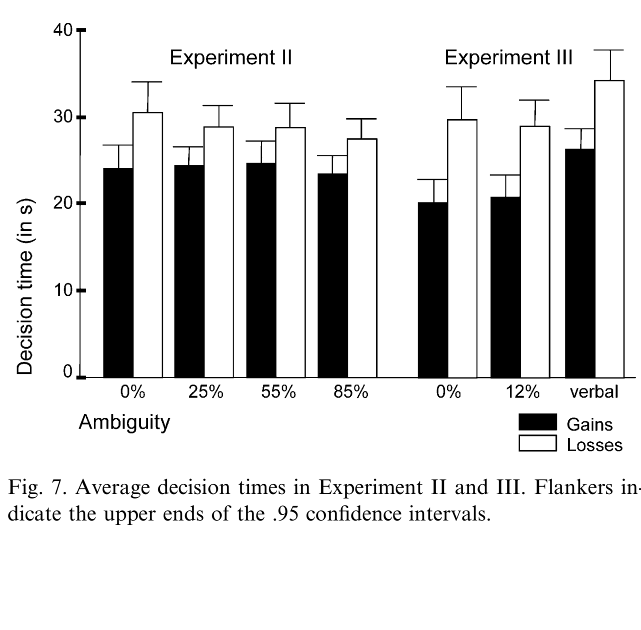

Fig. 7. Average decision times in Experiment II and III. Flankers indicate the upper ends of the .95 confidence intervals.

test). Single comparisons conducted for gains, following Conover (1980), found only the difference between missing and verbal ambiguity to be significant at the .05 level. Which of the three lists of probability phrases was used proved to be inconsequential. While ambiguity seemed to have induced careful behavior in the previous experiments, it now provoked risk seeking, and again, the results did not support the weighting hypothesis (θ did not decrease in the negative domain). One can only speculate why this experiment differed from the previous experiments. A difference in samples might be the reason because it does not seem to matter, whether ambiguity is achieved graphically or verbally. The somewhat stronger effects in the verbal condition might result from greater ambiguity: Although both conditions were intended to match with respect to ambiguity, 82% of the dms had felt better informed by the occluded spinners than by the verbal descriptions.

In Experiment III, too, decision times had been recorded. They were treated as in Experiment II and subjected to an ANOVA using the two repeated measurement factors *type/degree of ambiguity* and *gains/losses*. The first factor showed a clear effect ($F(2,51) = 21.7,\ p < .001$). From the right of Fig. 7, one can observe that this effect was solely caused by dms' taking on average 5.5 s (22%) longer for a verbal set. The second variable showed an even stronger effect ($F(1,52) = 75.0,\ p < .001$)—dms needed on average 8.6 s (39%) more to solve a negative set. No interaction occurred ($F(2,51) = .4, p = .66$). Again an increase in ambiguity within the graphical display did not prolong decision times as one might expect with venture theory.

3.2. Discussion

None of the three experiments reported here produced evidence in favor of the weighting hypothesis. This result seems to contradict another finding of González-Vallejo et al. (1994). As already mentioned, these authors supposed that verbal probability information would lead dms to increasingly base their decisions on lottery values. Consequently, dms should perform better in the verbal format when the expected values of the lotteries in a set depend predominantly on lottery values, i.e., $V(rev, r\$)$ is high. Yet, they should perform better in the numerical format when the expected values of the lotteries of a set depend mostly on lottery probabilities, i.e., when $V(rev, rp)$ is high. Indeed, the authors found this tendency in their data. However, this result does not speak unequivocally in favor of the weighting hypothesis either. The reason being that whenever dms are verbally informed about probabilities, some error comes into play that is not present within the numerical condition: This error stems from the fact that the same probability phrases have quite different meanings to different people (e.g., Mosteller & Youtz, 1990). And it is clear that the misjudgment of probabilities must hurt more in those sets in which probability information is especially important, i.e., when $V(rev, rp)$ is high.

All θ estimates reported here were based on a specific assumption about how dms evaluated lotteries. Such an assumption can be questioned. For example do Unequations (2) and (3) not take into account vagueness avoidance which for graphical ambiguity could be modeled such that lottery (a) is preferred over lottery (b) if and only if

$$\$_a(p_{\text{visible a}} + w\ ambig) > \$_b(p_{\text{visible b}} + w\ ambig), \qquad (4)$$

where w is a free parameter that can indicate vagueness aversion ($w < .5$) or seeking ($w > .5$). Lowering w has similar effects as increasing θ in Unequation 2.[4] Therefore, what has been presented here as a weighting phenomenon could similarly be presented as a vagueness avoidance phenomenon. However, at least in the experimental conditions of graphically achieved ambiguity it seems reasonable to me to neglect attitude towards vagueness because as all lotteries in a set showed the same degree of ambiguity this could neither be sought nor avoided. And in turn, the current design that fixes the level of imprecision in each set does not allow to test for effects of attitude towards vagueness.

I cannot prove that the derived θ estimates reflect the "true" weighting behavior of the participants. But they certainly corroborate that it would be premature to interpret the outcome prominence effect as a result of a change of strategy as implied by the weighting hypothesis.

I could find no evidence that people consider probability information to be less important when the precision of probability information decreases. Such behavior seems to be discordant with common sense and is very likely to hamper performance. How much does this lack of insight cost? And which weighting behavior is optimal when probabilities are ambiguous? It is the objective of the fourth part to answer these questions.

4. Evaluating decision strategies

This shall be done with the help of a computer simulation. In this simulation, different decision strategies were tested under a variety of circumstances. One strategy simulated the behavior identified in the experiments, other strategies were more rational. Employing different weighting behaviors enable identification of which weighting strategy works best and how costly it is

[4] $\$(p_{\text{visible}} + w\ ambig) = \$p_{\text{visible}} + w\ ambig\$$. Thus an increase of w increases the importance of the outcomes and the same holds for a decrease of θ in Unequation 2.

to depart from the path of virtue. Computer simulation not only enables one to directly compare different decision strategies but also allows one to take more ancillary conditions concerning the structure of the problems to be solved into account than any experiment could. For the sake of simplicity this second part will look only at the domain of gains.

4.1. Emulating participants behavior

To be able to test actual dms' behavior it is first necessary to emulate it. Two steps into this direction have already been taken: Firstly, Unequation 2 seems to be suited to describe how participants dissolve ambiguity and how they regard lottery values. And secondly, we know about average θ values. Two reasons call for the additional incorporation of a random error term into an adequate model of participants' behavior: (i) It was not possible to reproduce dms' rankings perfectly. And (ii) none of the participants reached the performance that would have resulted from strictly following the strategy Unequation 2 (see Hönekopp, 2000, for details). The actual dm behavior was modeled such that lottery (a) was preferred over lottery (b) if and only if

$$\begin{aligned} &error_{\mathrm{a}} \cdot \$_{\mathrm{a}} \cdot (p_{\mathrm{visible\_a}} + ambig/2)^{\theta} \\ &\quad > error_{\mathrm{b}} \cdot \$_{\mathrm{b}}(p_{\mathrm{visible\_b}} + ambig/2)^{\theta}, \end{aligned} \tag{5}$$

where $error_{\mathrm{a}}$ and $error_{\mathrm{b}}$ are uniformly distributed random variables with an expected value of 1. I shall refer to the depicted model as *empiric*. A computer simulation was used to determine the boundaries of *error* that best allowed reproduction of the pooled results of Experiments I and II. In this simulation, artificial dms that decided in accordance with *empiric* solved tasks typical of the first two experiments. The "dms" applied θ values such that the empirical distributions were reproduced. After the "dms" had finished, θ values were estimated in the same way as had been for the actual participants. In this way the artificial dms provided not only performance data, but the model fit could also be obtained and compared to the fitting of the actual data. On all four ambiguity levels, *empiric* could reproduce the empirical values very well when *error* was restricted to the interval [.45, 1.55]. Therefore, this specification was used as a model of dms' actual behavior.

4.2. The computer simulation—considered variables and design

In the next section, I shall describe the computer simulation in detail. All lotteries employed were randomly constructed, thereby following the guidelines described for Experiment I if not mentioned otherwise. In the following paragraphs I shall describe the variables that the simulation took into account.

Two different decision making under uncertainty tasks were examined. The *ranking paradigm* was the same as in the experiments. In this paradigm all possible courses of action were pursued, but with different force. Many natural decision tasks share this structure. We might, for instance, consider a manager who is responsible for the development of new products. She probably will not put all her eggs into one basket and give all resources to the best idea, but will promote several ideas with different force instead. In contrast the *choice paradigm* demanded that one choose one of several uncertain prospects. That means, one of the lotteries in a set had to be selected, and only this lottery was played. This task parallels situations in which only one course of action can be taken and all others are neglected. We might think of a patient who ponders over going to a medical check up or skipping it, or we may imagine members of a board who have to decide whether or not to buy another company.

Of course variation of the level of ambiguity was a key factor in the simulation. *ambig* varied from 0 to .9 in steps of .1.

It is desirable to consider the variability of \$ as well. This will change as the interval $[\$_{\mathrm{min}}, \$_{\mathrm{max}}]$, from which the values of the lotteries are drawn changes. Along with this the relative importance that the outcomes have on the lotteries' expected values will vary. $\$_{\mathrm{min}}$ was always 3, $\$_{\mathrm{max}}$ varied and took the values 3, 8, 11, 32, 47, 67, 97, 135, 190, 262, 365, and 670. This included the extreme case of no variability in the lottery values.

The impact of different decision strategies might change with the number of alternatives to be considered. Therefore, the simulation took into account several set sizes: 2, 3, 4, 6, 8, and 10 lotteries.

As already mentioned, *empiric* was not the only decision strategy scrutinized in the simulation. To have a yardstick against which to measure *empiric*'s performance, I additionally employed a strategy that resolved ambiguity in the optimal way. Following this strategy, *smart*, lottery (a) is preferred over lottery (b) if and only if

$$\$_{\mathrm{a}} \cdot p^{\theta}_{\mathrm{expect\_a}} > \$_{\mathrm{b}} \cdot p^{\theta}_{\mathrm{expect\_b}}, \tag{6}$$

where $p_{\mathrm{expect\_a}}$ and $p_{\mathrm{expect\_b}}$ are the expected values of the respective lottery's chances. If there exists any simple relationship between p_{expect} on the one hand and *ambig* and *p* on the other hand it escaped my inquiries, which is why I can give no such account here. However, the specific values, which where computed one by one, can be found in Hönekopp's (2000) Appendix.

In addition to *empiric* and *smart* a third strategy that is captured by Unequation 2 was used. I shall call this strategy *simple*. *Simple* is identical with *empiric* with exception of the error term in the latter. Therefore, the inclusion of *simple* into the simulation enables the examination of how much of the performance difference between *smart* and *empiric* is due to the latter dissolving ambiguity in a non-optimal manner or due to the influence of *error*.

To determine the optimal weighting behavior and the costs of deviations from it, nine levels of θ were employed: .33, .4, .5, .67, 1, 1.5, 2, 2.5, and 3. These values cover the whole range of values found in the experiments.

All factor combinations were conducted, resulting in a $2 \times 10 \times 12 \times 6 \times 3 \times 9$ factorial design (decision task, degree of ambiguity, variability of \$, set size, basic strategy, and θ). For each of the 38,880 resulting cells 50,000 sets of randomly constructed lotteries were analyzed.

4.3. Results

To allow for meaningful comparisons across conditions that differed systematically in outcomes (e.g., the ranking task yielded much more than the choice task), data were first standardized such that in each relevant subclass, the result of the best strategy was set at 100%. That means, 100% always denotes the result of optimal behavior under optimal conditions (no ambiguity). The values of all other cells were transformed into percentages accordingly.

I refrained from performing inferential tests because all effects of any size that matters can be considered to be statistically significant a priori, owing to the simulation's large *n*.

Three of the variables considered (degree of ambiguity, basic strategy, and θ) should be thought of as directly affecting performance. Ambiguity, for example, will necessarily derogate performance. In contrast, it does not make sense to think that set size or the type of decision task affect performance on their own. Instead one should consider these variables as mediating the influence of the three causal variables. The effects of the causal variables are depicted in Fig. 8. Here results are split for θ, basic strategy, and level of ambiguity. That is, each data point represents the average of 144 cells.

The central findings can be summarized as follows: (i) While *smart* and *simple* produced virtually identical results *empiric* fell on average 4.2% behind *smart*. Thus, investing much effort into optimally dissolving ambiguity obviously did not pay. But a lack of consequence, as introduced by *error* into *empiric*, proved to be comparably costly. (ii) Surprisingly, lowering θ with increasing vagueness was not optimal. *Smart* always performed best with a fixed θ of 1; *empiric* even required to *raise* θ with increasing ambiguity to achieve optimal performance. Thus, the (slight) tendency to increase θ with rising ambiguity, shown by dms in the Experiments I and II, is not ill-advised, but adaptive. Overall, however, weighting strategy had only a small impact on performance. (iii) Up to a middle level ambiguity derogated surprisingly little decision outcomes. For example, an ambiguity of .4 reduced *smart*'s average performance by only 1.8%, and reduced that of *empiric* by only 2.5%. In general, the same increase in ambiguity hurt more when it started at a higher level.

All effects were considerably stronger in the choice paradigm than in the ranking paradigm; on average, losses were 2.5 times higher in the former task. Along similar lines, average losses were 2.1 times higher in large sets (10 lotteries) as compared to small sets (2 lotteries). The impact of the relative variability of \$ proved to be negligible. No considerable interaction effects between the moderating and the causing variables occurred.

4.4. Generalizability of results

Ambiguity was not only a cornerstone of the simulation; it also proved to be the force with the strongest impact on performance. Being well defined and having neat boundaries, the simulation's ambiguity was of a type that will be hard to find outside the lab. Therefore, it would be desirable if the simulation's results could be

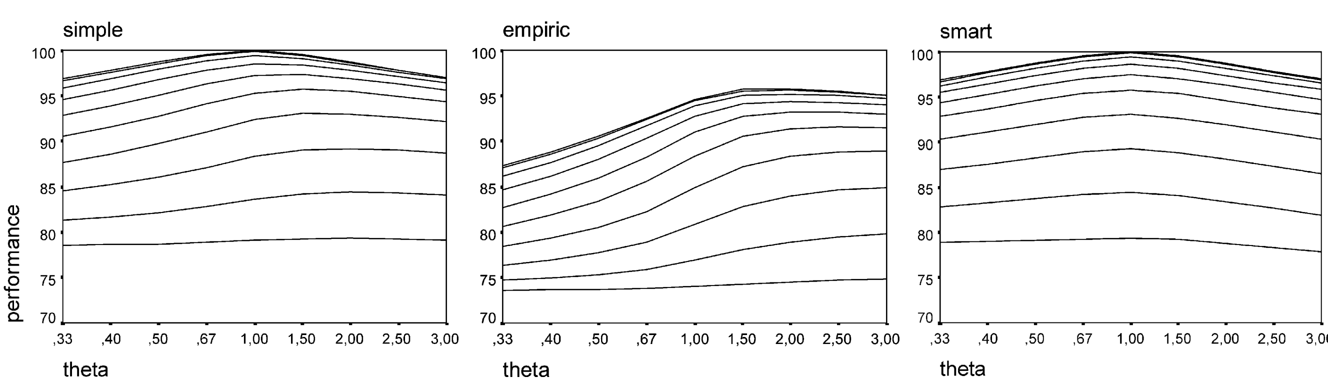

Fig. 8. Average performance of the three strategies for different levels of θ and ambiguity. The lowest line in each of the three sections marks an ambiguity of .9, the second lowest a level of .8 and so on. Each line constituting point represents the average of 144 cells.

generalized to more natural situations. Fortunately, at least one proper test is possible due to the use of verbal probability descriptions in Experiment III. If the simulation could correctly predict the costs of communicating probabilities verbally, this would give credit to simulation's generalizability (remember that *empiric* was designed independently of the results of Experiment III).

For such a test it is important to bear in mind that expressions like *likely*, *quite probable* and so on are not only vague, but furthermore, that different people have quite different ideas about what those terms mean (e.g., Mosteller & Youtz, 1990). Additionally, senders and receivers seem to differ systematically in their understanding of probability expressions (Fillenbaum et al., 1991). Therefore, the damage created by using verbal expressions for communicating probabilities is not restricted to their ambiguity, but some additional error is likely to come into play. However, the simulation should be able to grasp this aspect as well because simulation's *ambig* does nothing else but introduce some error into the perception of true probabilities. If this error does not stem from ambiguity, but from some other source (we might not only think of different interindividual meanings of verbal probability descriptions, but also simply of a dm who is misinformed by some expert about some chance relevant to the decision), it can be handled by converting this error into simulation's currency *ambig*.

Remember that in the critical condition of Experiment III, dms were informed about lotteries' probabilities by verbal translations. At the end of the experiment, dms retranslated these phrases into numbers. Thus, the information loss that was caused by the use of verbal descriptions can be measured. I used Tucker's congruence coefficient (Zegers & ten Berge, 1985) between true and retranslated probabilities as a measure of this loss. The median congruence proved to be $p = .983$. In the simulation's paradigm, the same congruence between true and perceived probabilities arises with an ambiguity of .34.[5]

For the condition of verbal probability information the simulation predicts that dms achieve 99.1% of the result that they yield in the condition without ambiguity.[6] In fact, performance was 98.2%, which is quite close to the expected value and not significantly different from it (one sample t test: $t = 1.1, df = 52, p = .27$). That is, a first test backs up the idea that the simulation's results can predict the effect of probability misjudgment arising from some mechanism other than the one specified in the simulation.

4.5. What can be learned from the simulation?

The simulation gives a clear-cut overall picture. Decision making performance was hard to disturb in the two tasks examined, and it showed to be quite robust against manifold nuisances: Non-optimal dissolving of ambiguity and non-optimal weighting showed hardly any effect; an already substantial level of ambiguity of .4 decreased *empiric*'s performance by less than 4% in the especially susceptible choice task; and even a serious flaw in decision strategy, as expressed by *error* in *empiric*, cost, on average, no more than 4%.

Such a general robustness has previously been demonstrated for other kinds of decisions: Gigerenzer and Goldstein (1996), concerned with comparative judgements of city populations, have found that inferences might not suffer or might even be improved by lack of knowledge. Likewise, the relinquishment of elaborate inference strategies did not do much harm: A very simple algorithm, which used for any comparison only the most valid cue, did as well as a complex regression model and different tallying strategies (see also Czerlinski, Gigerenzer, & Goldstein, 1999). Dawes and Corrigan (1974) have analyzed the predictive power of linear models for different fields and they have shown that a certain deviation from optimal weights altered performance only slightly. Even simple unit weighting of the predictors (i.e., predictors differ only in sign) yielded results that are comparable to those which stemmed from optimal models.

Drop in performance increased exponentially with level of ambiguity (see Fig. 8). This has an important implication for decision making policy in general: We should expect that the marginal return of any effort to decrease ambiguity (or generally speaking, of any endeavor to improve the calibration of dms subjective probabilities) will steeply decline. First, because a reduction from high to medium ambiguity is more effective than one from a medium level to a low one. And additionally, because the same reduction of ambiguity can be expected to be less costly when it starts from a high level of imprecision as compared to beginning at a lower level of ambiguity.

Participants' behavior, as modeled in the simulation, demonstrated three flaws: imperfect dissolution of ambiguity, non-optimal weighting behavior, and lack of strategic consequence (as expressed by *error*). The results of the simulation clearly show that any attempt to improve decision quality should focus on the last point; that is, performance should benefit most when dms learn to stick to a simple strategy.

The fact that the meanings of verbal probability descriptions vary considerably between people led several authors to criticize the common practice of communicating degrees of belief in this way. Beyth-Marom (1982) argued for using numbers instead or ranges, if numbers

[5] Result of a Monte Carlo simulation.

[6] This prediction was interpolated from the simulation conditions 30% and 40% ambiguity and it was based on the assumption, that dms in Experiment III would use a θ of 1.5 in the verbal condition as well as in the condition without ambiguity. This value was estimated from the pooled results of Experiments I and II.

would feign an unjustified degree of precision. Mosteller and Youtz (1990) proposed a codification of the meaning of verbal descriptions. The results of the simulation suggest that such measures would be of little help. In Experiment III, the use of verbal descriptions caused a misjudgment of probabilities as would be caused in the simulation by an ambiguity level of approximately .3; this decreased *empiric*'s average performance by not more than 1.4%. Note that verbal communication impairs decision only then by that much if the sender could alternatively state the true probability precisely. This will rarely be the case and, therefore, the disadvantage of verbal communication should even be smaller most of the time.

There are a lot of reasons that speak against a codification of probability phrases. Among them are impracticability (e.g., Wallsten & Budescu, 1990) and the dependency of the meaning of these phrases on context (Clark, 1990). Furthermore, Teigen and Brun (1995) showed that probability phrases carry much more meaning than a range of probabilities. One more reason against an extreme measure like a codification of phrases is that the improvement in precision of communication, which it might show, would only very slightly improve decisions depending on this information. But we should not forget that even though the *relative* price of verbal probability description is small, its *absolute* costs might still be considerable with high stakes. Therefore, a simple and cheap measure like the use of numbers might be considered valuable in many circumstances. However, I feel skeptical of Beyth-Marom's (1982) suggestion that senders should state a probability range instead of a single number so that receivers would be informed about the degree of precision of the probability estimate as well. Such a recommendation appears sound at first sight; however, the simulation suggests that such a policy might compromise subsequent decisions instead of improving them. How can a dm respond to the information's degree of precision? Only by adjusting her weighting behavior. But, as shown, to meddle with θ is not very promising. And any effort that dms directed towards optimal weighting might divert them from their key task: to stick staunchly to a reasonable strategy.

Although the simulation tried to cover a broad variety of circumstances, its potential to do so was necessarily limited. Therefore, it is easy to think of additional factors that might have changed the picture or that would have represented natural decision situations more adequately. Each reader will have something different in mind here. In the author's mind prevails the thought that the simulation would benefit a lot if it allowed for lotteries with different levels of ambiguity in the same set; this would come closer to everyday decision settings (however, to do so would require a much richer emulation of dms' behavior, because people often show a preference for low ambiguity). Therefore, the current simulation might be seen as a beginning and not as an end.

Acknowledgments

Part of this research was made possible due to a stipend of Westfälische Wilhelms-Universität Münster to the author. I would like to thank an anonymous reviewer, David Budescu, Wolfgang Hell, Romana Podgorsek, Peter Sedlmeier, and Elke Weber for their valuable contributions.

References

Beyth-Marom, R. (1982). How probable is probable? A numerical translation of verbal probability expressions. *Journal of Forecasting, 1*, 257–269.

Budescu, D. V., Weinberg, S., & Wallsten, T. (1988). Decisions based on numerically and verbally expressed uncertainties. *Journal of Experimental Psychology: Human Perception and Performance, 14*, 281–294.

Budescu, D. V., & Weiss, W. (1987). Reflection of transitive and intransitive preferences: A test of prospect theory. *Organizational Behavior and Human Decision Processes, 39*, 184–202.

Camerer, C., & Weber, M. (1992). Recent developments in modeling preferences: Uncertainty and ambiguity. *Journal of Risk and Uncertainty, 5*, 325–370.

Clark, H. H. (1990). Comment on Mosteller, F., & Youtz, C. (1990). *Statistical Science*, 5, 12–16.

Conover, W. J. (1980). *Practical Nonparametric Statistics*. New York: Wiley. Cited in J. Bortz, G. A. Lienert, & K. Boehnke. (1990). *Verteilungsfreie Methoden in der Biostatistik*. Berlin: Springer.

Czerlinski, J., Gigerenzer, G., & Goldstein, D. G. (1999). How good are simple heuristics? In G. Gigerenzer, P. M. Todd, and the ABC Research Group (Eds.), *Simple heuristics that make us smart*. New York: Oxford University Press.

Dawes, R. M., & Corrigan, B. (1974). Linear models in decision making. *Psychological Bulletin, 81*, 95–107.

Erev, I., Bornstein, G., & Wallsten, T. S. (1993). The negative effect of probability assessment on decision quality. *Organizational Behavior and Human Decision Processes, 55*, 78–94.

Erev, I., & Cohen, B. L. (1990). Verbal versus numerical probabilities: Efficiency, biases, and the preference paradox. *Organizational Behavior and Human Decision Processes, 45*, 1–18.

Erev, I., & Wallsten, T. S. (1993). The effect of explicit probabilities on decision weights and on the reflection effect. *Journal of Behavioral Decision Making, 6*, 221–241.

Fillenbaum, S., Wallsten, T. S., Cohen, B. L., & Cox, J. A. (1991). Some effects of vocabulary and communication task on the understanding and use of vague probability expressions. *American Journal of Psychology, 104*, 55–60.

Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review, 103*, 650–669.

González-Vallejo, C. G., Erev, I., & Wallsten, T. S. (1994). Do decision quality and preference order depend on whether probabilities are verbal or numerical. *American Journal of Psychology, 107*, 157–172.

González-Vallejo, C., & Wallsten, T. S. (1992). Effects of probability mode on preference reversal. *Journal of Experimental Psychology: Learning, Memory and Cognition, 18*, 855–864.

Hönekopp, J. (2000). Die Gewichtung von Wahrscheinlichkeitsinformation in Abhängigkeit von deren Genauigkeit bei Entscheidungen unter Unsicherheit. Ph.D. Thesis, Universität Paderborn. http://ubdata.uni-paderborn.de/ediss/02/2000/hoenekop.

Hogarth, R. M., & Einhorn, H. J. (1990). Venture theory: A model of decision weights. *Management Science, 36*, 780–803.

Kahneman, D., & Tversky, A. (1979). Prospect Theory: An analysis of decision under risk. *Econometrica, 47*, 263–291.

Lichtenstein, S., & Slovic, P. (1971). Reversals of preference between bids and choices in gambling decisions. *Journal of Experimental Psychology, 89*, 46–55.

Lopes, L. L. (1995). On modeling risky choice: Why reasons matter. In: J. -P. Caverni, M. Bar Hillel, F. H. Barron & H. Jungermann (Hrg.), *Contributions to decision making*. Amsterdam: Elsevier.

Mosteller, F., & Youtz, C. (1990). Quantifying probabilistic expressions. *Statistical Science, 5*, 2–12.

Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin, 95*, 109–133.

Page, E. B. (1963). Ordered hypotheses for multiple treatments. A significance test for linear ranks. *Journal of the American Statistical Association*, 58, 216–230. Cited in J. Bortz, G. A. Lienert, & K. Boehnke. (1990). *Verteilungsfreie Methoden in der Biostatistik*. Berlin: Springer.

Reagan, T. R., Mosteller, F., & Youtz, C. (1989). Quantitative meanings of verbal probability expression. *Journal of Applied Psychology, 74*, 433–442.

Rosenthal, R., Rosnow, R. L., & Rubin, D. B. (1999). *Contrasts and effect sizes in behavioral research. A correlational approach*. Cambridge: Cambridge University Press.

Savage, L. J. (1954). *The Foundations of Statistics*. New York: Wiley.

Svenson, K., & Karlsson, G. (1986). Attractiveness of decision alternatives characterized by numerical and non-numerical information. *Scandinavian Journal of Psychology, 27*, 74–84.

Teigen, K. H., & Brun, W. (1995). Yes, but it is uncertain: Direction and communicative intention of verbal probabilistic terms. *Acta Psychologica, 88*, 233–258.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science, 211*, 453–458.

Tversky, A., & Fox, C. R. (1995). Weighing risk and uncertainty. *Psychological Review, 102*, 269–283.

Tversky, A., Sattath, S., & Slovic, P. (1988). Contingent weighting in judgment and choice. *Psychological Review, 95*, 371–384.

Wallsten, T. S. (1971). Subjectively expected utility theory and subjects' probability estimates: Use of measurement-free techniques. *Journal of Experimental Psychology, 88*, 31–40.

Wallsten, T. S., & Budescu, D.V. (1990). Comment on Mosteller, F., & Youtz, C. (1990). *Statistical Science*, 5, 23–26.

Wallsten, T. S., Budescu, D. V., & Erev, I. (1988). Understanding and using linguistic uncertainties. *Acta Psychologica, 68*, 39–52.

Wallsten, T. S., Budescu, D. V., Rapoport, A., Zwick, R., & Forsyth, B. (1986). Measuring the vague meanings of probability terms. *Journal of Experimental Psychology: General, 115*, 348–365.

Wallsten, T. S., Budescu, D. V., & Tsao, C. J. (1997). Combining linguistic probabilities. In: Scholz, R. W. & Zimmer, A. (Hrg.), *Qualitative aspects of decision making*. Lengerich: Pabst.

Wallsten, T. S., Budescu, D., Zwick, R., & Kemp, S. M. (1993). Preferences and reasons for communicating probabilistic information in verbal or numerical terms. *Bulletin of the Psychonomic Society, 31*, 135–138.

Weber, E. U. (1994). From subjective probabilities to decision weights: The effect of asymmetric loss functions on the evaluation of uncertain outcomes and events. *Psychological Bulletin, 115*, 228–242.

Zegers, F. E., & ten Berge, J. M. F. (1985). A family of association coefficients for metric scales. *Psychometrika, 50*, 17–24.

Further reading

Lopes, L. L. (1995). On modeling risky choice: Why reasons matter. In: J. -P. Caverni, M. Bar Hillel, F. H. Barron & H. Jungermann (Hrg.), *Contributions to decision making*. Amsterdam: Elsevier.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science, 211*, 453–458.

Tversky, A., & Fox, C. R. (1995). Weighing risk and uncertainty. *Psychological Review, 102*, 269–283.

Wallsten, T. S., Budescu, D. V., & Erev, I. (1988). Understanding and using linguistic uncertainties. *Acta Psychologica, 68*, 39–52.

Received 27 June 2001